



PCT/AU00/00817

4

REC'D 15 AUG 2000

WIPO

PCT

AU00/817

Patent Office
Canberra

I, LEANNE MYNOTT, ACTING MANAGER PATENT ADMINISTRATION hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 1459 for a patent by JAMES QUEST filed on 06 July 1999.

WITNESS my hand this
Nineteenth day of July 2000

LEANNE MYNOTT
ACTING MANAGER PATENT
ADMINISTRATION

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



-1-

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

APPLICANT: JAMES QUEST
NUMBER:
FILING DATE:

Invention Title: SPEECH RECOGNITION SYSTEM AND METHOD

The invention is described in the following statement:-

communication system composed of arbitrary symbols which possess an agreed-upon significance within a community, are independent of immediate context and are connected in rule-governing ways.

5 Voice recognition techniques and technology (VRT) are well known and with the power of modern computing have advanced significantly in recent times. VRT has many significant applications including speech to text conversion for sound activated word processing, natural
10 speech synthesis for messages (the so-called Talking Timetable), voice activated hands-free control such as for example can be used for vehicular and appliance control, voice activated control of bio-medical devices for the disabled, name dialling for telephones etc.

15 However even with modern computing systems, the development of VRT has lagged its potential, particularly in relation to SE.

 It is believed by the inventor that VRT associated with SE has not achieved its potential and advanced to
20 the extent that technology would otherwise permit because modern SE is traditionally regarded as a stress timed language and is analysed for VRT purposes in accordance with the classical Saussurean universal language sign and the Applied Linguistic general definition and other
25 theories of language that are based upon them.

 This present invention is based on the inventor's understanding that rather than being a stress timed language, SE has an analytic (hereafter called analytical) phonology in which words, as defined, have
30 two values or orders of signification.

 The first value is a standardised or fixed phonetic value and identity which is set by convention, such for example, as defined by the phonetic entries of individual words found in dictionaries using the International
35 Phonetic Alphabet (IPA) and other systems of phonetic notation such as those used by American dictionaries.

 The second value is a variable phonetic value and identity. The second value is variable in time and sound

prominent or reduced. The expression "restricted syllable" refers to the syllable in a polysyllabic word that does not carry main stress, or main secondary or tertiary stress. In words in connected speech the
 5 restricted syllable may assume a value of stress equal to or less than the free and protected syllables in that word.

It is also preferred that the variable indicators include features of speech such as pitch, tone, duration,
 10 rhythm and tempo.

It is further preferred that the variable indicators include other suprasegmental or prosodic features of speech such as volume, speed of delivery, enunciation, pausing, phrasing and word linking.

15 In a preferred embodiment of the invention the method further includes:-

recording speech spoken by a speaker;

indicating to the speaker the meanings of the variable indicators of the recorded speech, and

20 designating or affirming the meanings of the variable indicators indicated to the speaker.

It is preferred that the method also includes storing data representative of analysed words for which the meanings of the variable indicators have been
 25 designated or affirmed.

In another aspect this invention resides broadly in a system for recognising speech consisting of words having syllables and phonemes, the system including:-

30 recording means for recording speech spoken by a speaker;

means for assigning a first order of signification to a word, the first order of signification including standardised indicators having agreed meanings independent of the speaker;

35 means for assigning a second order of signification to a word, the second order of signification including variable indicators having meanings which are generated by the speaker and are dependant on the context of the

It is preferred that the language is spoken English and that it is taught as a first language.

It is also preferred that the language is spoken English and that it is taught as a second language.

5

Description of Drawings

In order that this invention may be more easily understood and put into practical effect, reference will now be made to the accompanying drawings which illustrate
10 a preferred embodiment of the invention, wherein:-

FIGS 1A to 1F diagrammatically illustrate a second order signification of the word "disappointing" and which illustrates free, restricted and protected syllables;

FIGS 2A to 2G diagrammatically illustrate seven
15 commonly understood uses of pitch and tone in SE, providing various examples of their use and readily understood meanings;

FIGS 2H to 2J diagrammatically illustrate the uses of variable pitch and tone in spoken phrases to compress
20 additional meaning into speech in the second order of signification;

FIG 3 is a flow chart of a method of recognising speech in accordance with the invention, and

FIG 4 is a schematic block diagram illustrating a
25 system for recognising speech in accordance with the invention.

Description of Preferred Embodiments of Invention

Before providing a more detailed description of the
30 preferred embodiments of the methods and description of this invention, a description of the inventor's understanding of underlying principles will be given, first at a more general level and then in summary.

35 - Synthetic and Analytical languages

The key languages that helped to shape modern English are what linguists call synthetic languages. Modern German, for example, is a synthetic tongue, as was

governed by a stress timed phonology. The hallmark of the entire English language system, both written and spoken, is that it is an analytical system. Ergo, the connected speech of everyday English is governed by an analytical phonological system.

- Stress timed languages

Stress timed language systems include spoken German, Russian, Arabic and Greek. (SE has also traditionally been regarded, incorrectly in the view of the inventor, as a stress timed language). These spoken language systems are governed by a different timing principle than syllable timed phonologies.

The main tenets of a stress timed language are that:

(i) The function of main stress is to give prominence to certain words, or even one word, within an utterance, main stress falling on the salient syllable of the stressed word.

(ii) Main stress beats fall at roughly regular intervals of time within connected speech with weaker stress beats falling on the words and syllables in between. (In English Linguistics, the chunk of speech between two main stress beats is often called a foot, with feet said to be "about the same".)

(iii) Stress, therefore, produces a clear, underlying rhythm in streams of natural speech as it is organised around a stable timing principle.

(iv) Stress and rhythm, in their own right, are incapable of, or severely limited in, expressing contrasts in meaning, and

(v) The phonological system imposes such order from within.

- The Analytical Phonology

Because modern SE is not a stress timed language as are those Germanic tongues governed by a stress timed phonology, it is beneficial to consider SE as being represented by a new category of phonological system -

restricted syllable, the gradations of stress in speech are otherwise freely transferable between phonemes (particularly the vowel sounds), syllables and words within utterances.

5 There are many and various gradations of stress obtainable in natural connected speech, not merely weak and strong. In general, stress in SE cannot be regulated nor categorised according to any external or fixed measure of gradation that the phonological system
10 imposes.

 The gradations of stress that attend the phonemes, syllables, words and phrases are relative to the utterance, having due regard for the immediate context of the utterance within its particular stream of connected
15 speech. It is on this basis that the listener evaluates the gradations and contrasts of stress audible within the speaker's speech and judges what is being made prominent and salient by the speaker in terms of both sound and words.

20 Stress, as defined above, is capable of signifying meaning or facilitating it.

 The speaker may use stress to make certain syllables and words prominent in sound and meaning, but main stress does not always signify prominence as this can sometimes
25 be achieved by using weak stress, if weakly stressed syllables can create prominence by standing out in the stream of speech. Prominence may also be achieved by way of the other phonological and prosodic features of everyday connected speech. In making prominent certain
30 phonemes, syllables, words and phrases in speech the speaker may wish, for example, to highlight certain words, place special emphasis and focus on certain words and parts of speech, or juxtapose, rank, contrast, infer or complement meanings and ideas embodied within and
35 between utterances.

 The uses to which stress in everyday SE may be put are at the behest of the individual speaker, precisely for the purposes of generating and producing further

- Sound and Timing Variables in Spoken English

The key phonological functions of SE dependant on variable stress (hereafter called the key phonological functions or the key phonological features) include
 5 variable timing and duration, variable rhythm and tempo and variable pitch and tone. Other suprasegmental and prosodic features of speech effected by variable stress (hereafter called other suprasegmental features) include
 10 speed of delivery, volume, word linking/not linking, enunciation, pausing and phrasing.

All of the above key phonological features and phonological functions and the other suprasegmental features of everyday connected English speech are inextricably linked to variable stress. Stress is well
 15 understood to have a commanding role to perform in the organising functions of the phonological system. However because stress in SE is variable, the interconnected phonological functions are also variable.

Because stress is variable the role of the other
 20 suprasegmental features of speech becomes important. These features generally are variable and/or optional. Their effects on the phonological functions in connected speech greatly heighten the system's overall capacity and scope to accommodate greater sound and timing variations
 25 and contrasts which variable stress enables.

All of the sound functions and features of everyday speech are capable of variation within a wide and legitimate parameter of sound & timing variability that the English phonology permits and encourages and which,
 30 crucially, the English language phonetic system tolerates and accommodates.

The factors of sound and timing variability in connected speech work at the basic level of the phoneme, the syllable and the word and at the more general level
 35 of spoken phrases and connected speech.

Sound and timing variations effect or change the way the native speaker pronounces phonemes, syllables, words and phrases in ways the listener can hear.

creating prominence in some way?

. What are rhythmic and tempo changes signifying?

. What are pauses signifying, framing or announcing?

A fundamental axiom of the analytical phonology is
 5 that the native speaker is free to apply any of the many
 variable phonological functions and suprasegmental
 features in the pronunciation of phonemes, syllables and
 words at choice which are obtainable in speech for the
 purposes of creating further meaning provided that such
 10 variations do not negate the agreed-upon meanings of the
 words.

To these ends and within this fundamental rule,
 further meaning can be generated when the phonemes and
 syllables that constitute the word - still an arbitrary
 15 symbol possessing a static agreed-upon meaning - are
 varied in sound and timing according to the speaker's
 pronunciation and in ways the listener hears and
 registers as meaningful.

Having regard to the speaker's manner and habits of
 20 pronunciation, the variation of sound and timing and the
 application of the other variable and optional
 suprasegmental features at the level of the phoneme,
 syllable and word, cannot but effect the structure,
 qualities and organisation of spoken phrases and larger
 25 passages of discourse as connected speech progresses.

Within phonemes, syllables, words and phrases, and
 within the general flow of connected speech, specific
 sound and timing variations (hereafter called variations)
 will combine to create discernible sound and timing
 30 contrasts (hereafter called contrasts). Therefore,
 in terms of the sound imagery of SE, the following
 equation applies:-

Phonemes, syllables and/or words + X number of
 variations = Words and phrases + Y number of
 35 appreciable contrasts in speech.

The potential number of variations and specific
 combinations of variations capable of creating
 discernible contrasts within connected English speech is

other utterances are within or without the immediate context.

This further meaning, which may be termed second order meaning, is communicated to the listener at the same time that the word's agreed-upon meaning is signified.

The agreed-upon meaning of the actual words which remains constant throughout may be referred to as the first order of meaning.

The legitimate variations and contrasts that can produce second order meanings not only create further meaning/s in the mind of the listener/s but may also create new meaning/s in the mind of the listener/s

Since the particular combination of sound and timing variations that might create a particular contrast, or cluster of contrasts, cannot be pre-defined - as this is something that is initiated and articulated by the individual speaker in the moment of utterance and within the extemporaneous circumstances of the particular context - we are unable to predict every possible context nor classify the mood, manner of speech and temper of mind of every individual speaker.

The onus then falls upon the listener to be able to recognise, register, decode and interpret variations and contrasts for second order meanings. This is something at which the fluent native speaker is proficient. In short, sound and timing variations and contrasts in the sound imagery, or soundscape, of everyday connected English language speech work productively to be the signifiers of second order meanings.

This system through which the production of second order meaning in SE occurs may be termed the second order of signification.

The traditional and orthodox linguistic system of producing sounds that signify static agreed-upon meaning, common to all languages, and as conceptualised by Saussure, may be termed the first order of signification.

tolerate marked and sometimes radical sound and timing variations within individual phonemes which having no one fixed principle of timing the system can permit. The phonetic system will indifferently
 5 accept legitimate variations without negating the static agreed-upon meaning of the word and without rendering the individually varied phoneme that constitutes the word incomprehensible to the listener.

10 This remarkable tolerance of the English language phonetic system is exemplified by the ability of all of the six written vowels a, e, i, o, u and the semi-vowel y to assume the reduced vowel sound /ə/ in certain and numerous syllables and
 15 words and in liberal application, particularly in free and restricted syllables.

It also enables any of the 44 sounds of SE, as defined by the International Phonetic Alphabet (IPA), to be stretched or elongated according to any
 20 timing format or aesthetic purpose that the individual speaker can successfully obtain in natural speech. This function particularly effects the vowel sounds of natural SE which are the phonemes most commonly varied.

25 The absence of any one fixed timing principle enables ordinary vowel sounds to become diphthongs, and diphthongs to become "triphthongs" and so on. The ability to elongate sounds, vowels in particular, is restricted only by the respiratory
 30 limits of the speaker and defined by the context of the utterance and according to the speaker's intention to create second order meaning of some kind.

Unlike the phonetic system of other language
 35 systems, the phonetic system of SE does not have a basic, fixed and limited stock of phonemes that are used to construct syllables and words. Instead the phonetic identity and the quantity of legitimately

two values in the way they are stressed.

The first "word stressing" is standardised by convention and defined in the phonetic entries of individual words found in dictionaries which usually mark
5 which syllables within polysyllabic words customarily assume main, secondary and weak stress.

The second "word stressing" is a variable that is defined by the individual speaker in the moment of utterance and which is relative to the circumstances of
10 the word's immediate "real life" context and its place within the flow of connected speech.

The system of variable word stressing, as referred to above, operates within certain limits and parameters. These limits are defined by the three kinds of syllables
15 a polysyllabic word may obtain within the flow of everyday connected speech.

The three kinds of syllables are:-

(i) The distinct or protected syllable.

This is the syllable that according to the normal
20 "dictionary" standard carries the main stress of the word. When the protected syllable does not carry main stress within an utterance it may then assume any gradation of stress so long as it is pronounced as distinct as is necessary so as to not to disturb the
25 word's internal distribution of stress to the extent that the word's agreed-upon meaning becomes negated or unclear. Both syllables in compound nouns can be regarded as protected syllables.

(ii) The free syllable.

30 This syllable can assume any gradation of stress within an utterance, from being the most reduced sound to the one carrying main stress and any point in between. It is different from the protected syllable, as it is the syllable which dictionaries define as carrying secondary
35 stress of one level or another. In many cases within the context of the phrase it may carry more stress than the syllable/s that by normal definition should carry the main stress of the word. When this occurs, or in

protected syllables in the analytical phonology.

FIG 1A shows the standardised dictionary stress pattern for the polysyllabic word "disappointing" which consists of the four syllables: *dis*, *ap*, *point* and *ing*.

5 However the word is subject to much variation in the traffic of connected speech when placed in different phrases carrying different second order meanings.

Thus as illustrated in FIG 1B which shows the potential variation of syllables in connected speech.

10 . the first syllable *dis*, which dictionaries define as having secondary stress, is a free syllable. The last syllable, the inflection *ing* in this case, is also a free syllable;

 . the second syllable *ap*, carrying the reduced sound /ə/, is the restricted syllable, and

15 . the third syllable *point*, carrying main stress, is the protected syllable.

 In the phrase "The result was disappointing" illustrated in FIG 1C, the stressing of the word
20 "disappointing" is as the dictionary would prescribe.

 In the phrases "A disappointing result" seen in FIG 1D and "That's disappointing" seen in FIG 1E, the word "disappointing" is still prominent and is stressed variably according to the speaker's construction of the
25 phrase and the particular second order meaning he or she wishes to convey.

 In the phrase "Very disappointing indeed" seen in FIG 1E, the word "disappointing" does not obtain prominence in the phrase at all. The speaker chooses to
30 place the vocal emphasis on the words *VERY* and *indeed*. In all examples where the standardised stressing of the word "disappointing" has been varied it is done so in order to generate a second order meaning within the utterance so as to signify an emotional or subjective
35 content or meaning.

 A word will deviate from its standardised word stressing and assume a variable word stressing and variable gradations of stress and prominence in connected

of words but moreover serve as the signifiers of second order meaning.

In all cases the listener's and observer's knowledge of SE, the speaker/s and the immediate context will
5 enable such second order meanings to be understood.

Also listeners and observers rely on commonly understood indicators of pitch and tone widely used in connected speech as a reference point by which legitimate sound and timing variations and contrasts may be
10 recognised, registered, decoded and evaluated for meaning.

Reference will be made to FIGS 2A to 2G which illustrate some commonly used and readily understood uses of tone and pitch in connected speech.

15 FIG 2A shows the use of a rising pitch and tone, commonly understood to signify a yes/no question, a clarifying question, a request for repetition, and interested feedback. In general, rising pitch and tone at the end of phrases or words commonly signifies a sense
20 of doubt, incompleteness, or a need to know more on the part of the speaker.

FIG 2B shows the use of a low rising pitch and tone commonly employed by speakers when reading items from a list to signify that the list is not yet complete, and
25 commonly to signify more neutral feedback or mild interest in what is being said.

FIG 2C shows the use of flat or level pitch and tone in speech, commonly employed to signify disinterest, boredom or sarcasm. In general, level tones will also
30 commonly attend routine or impersonal conversational exchanges.

FIGS 2D and 2E show the use of rising-falling and falling-rising pitch and tones, respectively. Both tone patterns are commonly understood to signify greater
35 emotional content and expression attending the speaker's speech, or to signify contrasting or competing meanings, or to signify a change in register, mood or conversation topic on the part of the speaker.

regards the answer to his or her own question.

Declaratives are subject to the same effects of variable tone and pitch.

Reference will be made to FIGS 2F and 2G. Speaker
 5 A's utterance: I'm going to stop drinking, may express
 opposing ideas and meanings simultaneously. The words
 themselves, the phrase's first order of meaning, express
 an apparently self-confident resolution, particularly as
 the words 'going' and 'stop' are stressed. However, a
 10 rising tone and pitch at the end of the phrase around the
 word 'drinking' could betray the speaker's own sense of
 doubt or commitment to his or her stated intention,
 particularly noticeable if the phrase is strongly
 stressed.

15 Similarly, in FIG 2G Speaker B's response: I ' m
 sure you will, with the contours of intonation curling
 upwards on the last word pitches a statement of fact on
 an uncertain note, subtextually suggesting Speaker B's
 sense of doubt, mistrust or ambivalence concerning
 20 Speaker A's utterance.

Another example of this is the Australian tendency
 to finish statements on a rising tone (called the 'mid
 rising pitch') which combines two meanings through both
 first and second orders of signification. The speaker is
 25 telling the listener something, narrating some past
 event, or relating information, by way of the words and
 grammatical form of the first order of signification,
 while the voice is producing second order meaning: ie,
 the mid rising pitch and tone at the end of each phrase
 30 is embedding the declarative in an habitual yes/no
 question: signifying a kind of abiding doubt or
 tentativeness on the speaker's part or a need to
 constantly check with the listener that he or she
 understands and is engaged in the conversation.

35 A fluent speaker may also express a pejorative idea,
 a criticism, reproach or complaint by exploiting the
 opportunities the second order of signification offers.
 Rather than deliver an insult in actual words, the

tendency of the 'mid rising pitch' in natural speech, hitherto unknown in that part of the world.

Thus, subtextual content, emotion, irony, idiomatic codes, complementary moods and meanings can be compressed
 5 within the one spoken English phrase without need of the speaker formulating a new spoken phrase in order to express these further meanings.

In this way the analytical purpose of the spoken English language is greatly served.

10

- The Analytical Phonology and Phrases and Sentences

Within the parameters defined by the free, protected and restricted syllables, stress is freely transferable
 15 between phonemes and syllables within connected speech.

This enables any word within an utterance to obtain prominence, or to stand out in some way within the phrase in the flow of speech. In a particular word in a phrase being made prominent or noticeable in sound and meaning,
 20 while the other words in the phrase do not gain as much prominence or do not stand out in the same way, the agreed upon meanings of all the words in the phrase nonetheless remain constant.

For example, any word in the phrase "You are going!"
 25 may gain prominence or stand out in some fashion without negating the agreed upon meanings of any of the words in the phrase. Moreover if the speaker were to make a particular word in the phrase especially prominent or noticeable then the cumulative meaning of the utterance
 30 would immediately change. This is because the word that gains prominence or stands out in the flow of speech can signify second order meaning.

Thus in addition to the agreed upon meanings of the words in the phrase being signified second order meanings
 35 may, in the first instance, also be signified according to the speaker's arrangement and placement of stress within the phrase.

the sound and timing qualities of the speaker's sound imagery, especially the contours of pitch and tone and other extemporaneous suprasegmental features that attended the prominent or noticeable word:

5 "YOU are going!" in placing the focus on the subject could also express the speaker's subjective opinion of the subject (e.g, enthusiasm or disgust);

 "You ARE going!" in placing the attention on the subject's intention "to go" may also signify a
10 curt imperative and not simply be a casual affirmation, and

 "You are GOing!" in placing the attention on the subject's act of "going" may also express the speaker's personal feelings in regards to the
15 subject's act of 'going' (e.g. regret or relief).

 The flexibility of the phonetic and phonological systems of the analytical language enables the sound patterns of individual phonemes to vary greatly in speech, and within the parameters of the free, protected
20 and restricted syllables also enables the stressing of individual words to vary greatly in speech. This is because the phonetic pronunciation of words, and word stress must be subject to such variation when individual phonemes, syllables and words are placed in phrases and
25 sentences in order that the speaker has the vocal means by which to choose which particular words are to be prominent or noticeable within and between utterances in the flow of speech in order to signify the second order meanings that they wish to communicate. As well as this
30 the liberating effects variable stress have on the other phonological functions and the other suprasegmental features greatly assist the speaker in creating prominence in words and phrases in speech which can generate further or more complex second order meanings.

35 All this greatly furthers the cause of the analytical language system in expressing more meaning with reduced words, forms and structures.

meanings.

Within the language's second order of signification it can be said that:

5 in the speakers' use of sound and timing variations and contrasts, variable sound images of standard words are obtained in speech in a way that does not negate the first order signification but, moreover, systematically generates the signifiers of further or new meaning, ie the second order of meaning.

10 Within the second order of signification the basic signifiers in the production of meaning in speech are phonemes which construct spoken phrases which are the basic units of second order meaning.

Therefore, it can be said that words within the 15 analytical phonology of SE possess two values:

- . The first value, as the arbitrary symbol Saussure so defined, and
- . A second value that the word may obtain in speech as a subjective or relative symbol.

20 The word's second subjective or relative value emanates from its first value. The word's subjective or relative value is realised in speech by the speaker's systematic variations and contrasts of sound and timing which the analytical language permits in the 25 pronunciation of its phonemes, syllables, words and phrases.

Therefore, meanings produced within the second order of signification of the analytical phonology of SE are speaker generated and context dependent.

30

- Summary of underlying principles

Unlike conventional Indo-European languages the phonology of SE is analytical. Its core drive is to find ways to express as much meaning as possible in the 35 shortest forms, words, and grammatical structures it can.

SE cannot be accommodated within Saussure's concept of the universal language sign and Applied Linguistics' general definition of language because it is a different

phonemes, syllables, words and phrases can be greatly varied. In speech, this gives the word its second value.

Variation in the pronunciation of phonemes, syllables and words in connected speech is possible
5 because stress in the phonology of SE is variable, as it is fixed to no one stable or regulated principle of timing.

Variable stress in SE means the other phonological features of speech (pitch, tone, duration, rhythm and
10 tempo) are also variable as they are dependent upon stress.

Other suprasegmental and prosodic features of everyday speech (volume, speed of delivery, enunciation, pausing, phrasing, word linking or not linking) are also
15 freed up because SE stress is variable.

Because sound and timing variations and contrasts operate at the basic level of the phoneme, there is not a small and limited stock of phonemes in SE - as in most languages - that are used to construct intelligible
20 words. With the many legitimate variations, contrasts and combinations freely available to English speakers, it is difficult to reliably count how many phonemes there are.

Each speaker is free to vary phonemes (particularly
25 vowel sounds), syllables and words by the way he or she delivers and pronounces them with his or her voice.

Legitimate variations in the pronunciation of ordinary words do not negate the agreed-upon meaning of the words (the first order of signification) but moreover
30 simultaneously inject or compress more meaning into everyday spoken phrases (the second order of signification). Thus SE finds a way to express more meaning with reduced forms.

In terms of the variable stressing of polysyllabic
35 words in connected speech, three kinds of syllables exist: the protected syllable, restricted syllable and free syllable.

In SE, second order meanings are speaker generated

and context dependent it is difficult if not impossible to define every possible or potential second order meaning that may be generated or created as we cannot predict every possible context, nor classify the mood, manner of speech, and temper of mind of every individual speaker who might generate second order meanings in the manner of his or her pronunciation.

Thus new second order meanings in SE are constantly created. Second order signification lends the already huge formal vocabulary of the English language - its corpus of arbitrary symbols - a virtually inexhaustible reservoir of potential second order meanings as these arbitrary symbols expand to gain new versions of meaning as they find their place in the everyday contexts of animated speech within the analytical system.

Turning now to the preferred embodiments illustrated in FIGS 3 and 4, there is shown in FIG 3 a flow chart of a method of recognising speech in accordance with the invention.

The method of recognising SE (which consists of words having syllables and phonemes) includes assigning to SE words a first order of signification (12) which includes standardised indicators having agreed meanings independent of the speaker. It also includes designating the syllables in words as being protected, restricted or free syllables which assigns a potential variability to each syllable that may be obtained in connected speech. A second order of signification is also assigned to SE words (13), this second order having variable indicators which have meanings which are generated by the speaker and are dependant on the context of the word in the flow of connected speech.

These variable indicators include the pronunciation (14) of phonemes, syllables and words in the speech with the syllables categorised as being either free syllables (15), protected syllables (16) or restricted syllables (17) according to the syllables' variable indicators

Analysing means 37 analyses words or speech in accordance with the first and second orders of signification and data representative of the analysed words, for which the meanings of the variable indicators
5 have been designated or affirmed, is stored in storage means 38.

Assignment means 32 and 33, analysing means 37, storage means 38 and transforming means 39 are embodied in a suitably programmed computer 41, the peripherals to
10 which include recorder 31 such as a microphone, indicating means 34 (screen), designating means 36 (keyboard or mouse) and output means 40 (printer, screen, speaker etc).

It will be obvious to those skilled in the art that
15 there are numerous ways in which the present invention can be practised in VRT, that suitably skilled programmers can write software embodying the teachings of this invention in its various applications to VRT, and that suitably configured and programmed computing systems
20 can be utilised to practice the invention.

Whilst the invention in its application to VRT can be adequately performed by those skilled in the art on the basis of the description thus far provided, a number of features and aspects relating to implementation of the
25 invention are further provided. These are not to be construed as limiting on the scope of the invention.

With reference to the preferred VRT implementation of the present invention:-

. The system is designed for application at the
30 level of the individual speaker or the individual operator who will use the system.

. The system first defines the standardised value of words according to their phonetic content and agreed-upon meaning as is defined by a particular
35 community of native English language speakers. The individual speaker has the opportunity for direct input into defining or editing what the standardised "sound" values and meanings of arbitrary symbols are

using the reference point of the standardised sound values.

5 The system notes the individual's particular tendencies, habits and patternings of pronunciation and voice in natural spontaneous speech and then alerts the operator to sound variations and contrasts evident in the operator's samplings of natural connected speech (as compared to the standardised values of words already established).

10 The system then allows the individual speaker to define the particular second order meanings that such variations and contrasts may signify. For instance, do certain combinations of sound and timing variations (say the elongating or reducing of
15 the vowel sound in the same word when repeated) mean something? Or do they signify nothing? Do clusters of contrasts noted by the sampling process in the operator's natural speech and which cause individual phonemes, syllables and words to noticeably vary and deviate from the standardised norm, possess second
20 order meaning, and if so, what? An emotion, irony, a subtext of some kind? Often an individual speaker's peculiar habits and manner of pronunciation signify a personal meaning, mood or theme, irrespective of what the actual words in the
25 phrase might be, but are still relative to and dependent on the immediate context. In these instances the system offers the operator the choice of identifying and labelling such recurring
30 variations and contrasts for any specific meanings they may possess.

 The definition of second order meanings is preferably effected by systems of menus, sub-menus and options over which the trained operator has
35 executive control. Individual operators are allowed to define what sound and timing variations and contrasts in their speech signify or mean.

 In order to correctly define second order

second order signification present only in speech will not be lost or nullified when translated or put into the written word.

5 Allowing individual English language speakers who are hearing impaired or deaf to appreciate second order signification and second order meanings that are encoded within natural speech through the written word.

10 Allowing individual English language speakers who are sight impaired or blind to appreciate second order signification and second order meanings encoded within the written word and which can be translated back into natural speech.

15 The present invention can also be used to teach how to speak a language. In use this method includes:-

20 assigning first and second orders of signification to words in speech, the words having syllables and phonemes, the first order of signification including standardised indicators having agreed meanings independent of the speaker and the second order of signification including variable indicators having meanings which are generated by the speaker and are dependant on the context of the word in the flow of connected speech, and

25 practicing speaking using different variable indicators in the second order of signification.

30 When teaching how to speak the English language in schools in English speaking countries information concerning the analytical nature of the language should be included in mainstream English curriculums.

A number of aspects associated with the present invention are emphasised in the curriculums. These include:-

35 Providing and teaching a proper and appropriate explanation of the nature, practices and principles of the analytical phonology of spoken English.

 Explaining to children, from the moment they learn to read and write, the nature of the

enormous executive power delegated to the individual speaker in producing second order meanings according to the individualistic way they may vary the sound and timing qualities of words when used in everyday connected
5 speech.

Present VRT package programs require the operator to painstakingly sample speech word by word. This sees English words as purely arbitrary symbols with no regard to the words' potential second relative or subjective
10 value that is clearly obtainable in natural connected speech. Existing VRT systems record the word's static agreed-upon meaning in the first order of signification where words as arbitrary symbols possess no more than this and generally have a standardised pronunciation.

15 Consequently for the computer in existing VRT packages to familiarise itself with the particular operator and recognise the operator's manner of speech, the onus falls heavily on the operator not to vary his or her pronunciation of the sampled words in any substantial
20 way when the system is later in use, lest the machine be unable to recognise the words correctly.

The failure of known VRT systems to come to grips with the second order of signification is believed by the inventor to have prevented computer science from
25 developing the technologies to enable communication with native English speakers at a level above the somewhat robotic-like current speech level.

The invention is also applicable in telecommunications where existing recorded computer
30 generated voices repeat back numbers to the telephone customer. These can be modified to sound more like natural connected speech.

It will of course be realised that whilst the above has been given by way of an illustrative example of this
35 invention, all such and other modifications and variations hereto, as would be apparent to persons skilled in the art, are deemed to fall within the broad scope and ambit of this invention as is herein set forth.

4, wherein said syllables are categorised as being either free syllables, protected syllables or restricted syllables.

5 8. A method of recognising speech as claimed in claim 3 and including:-

recording speech spoken by a speaker;

indicating to the speaker the meanings of the variable indicators of the recorded speech, and

10 designating or affirming the meanings of the variable indicators indicated to the speaker.

9. A method of recognising speech as claimed in claim 8 and including:-

15 storing data representative of analysed words for which the meanings of the variable indicators have been designated or affirmed.

10. A system for recognising speech consisting of words having syllables and phonemes, said system including:-

20 recording means for recording speech spoken by a speaker;

means for assigning a first order of signification to a word, said first order of signification including
25 standardised indicators having agreed meanings independent of the speaker;

means for assigning a second order of signification to a word, said second order of signification including variable indicators having meanings which are generated
30 by the speaker and are dependant on the context of the word in the flow of connected speech;

indicating means for indicating to a speaker the meanings of the variable indicators of the recorded speech, and

35 designation means whereby a speaker designates or affirms the meanings of the variable indicators indicated by the indicating means.

for the purposes of recognising and evaluating the speech for meaning.

5 16. A method of teaching how to speak a language as claimed in claim 13, wherein said language is spoken English.

10 17. A method of teaching how to speak a language as claimed in 16, wherein English is taught as a first language.

15 18. A method of teaching how to speak a language as claimed in 16, wherein English is taught as a second language.

JAMES DAVID QUEST

by

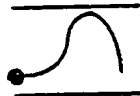
PIZZEYS PATENT AND TRADE MARK ATTORNEYS

RisingA yes or no question: *Are you hungry? Now?*A clarifying question: *What happened then?*Please repeat: *Sorry? Pardon? Again?*Involved feedback: *Really?*FIG 2ALow rising

Reading items from a list: more to come

*Sydney, Canberra...*Neutral feedback: *Uh-huh, Mmmm*FIG 2BLevel

Monotone.

Bored or indifferent: *Hello.*Ironic, sarcastic: *Sorry. Really?*Routine: *Next please.*FIG 2CRising-fallingExclamatory: *Fabulous! Awesome!*Opinion: *Well I think...*Dramatic: *Now!*Conversation markers: *Anyway...*FIG 2D**KEY***free syllables***restricted syllables****protected syllables**

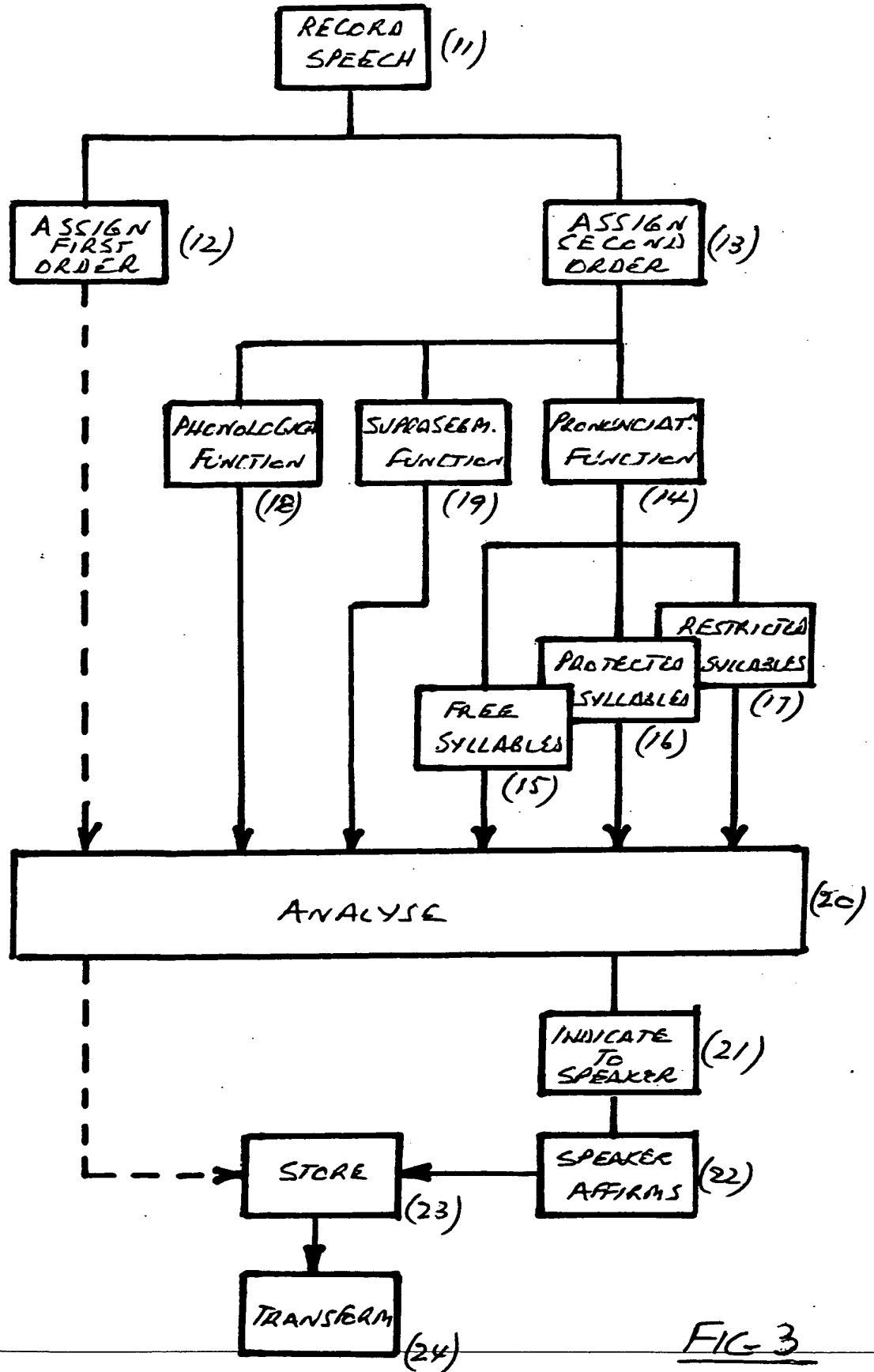


FIG 3